# Subclinical Narcissistic Traits in the Age of Algorithmic Validation:

## A Critical Review of Reinforcement Mechanisms and Attachment Vulnerabilities

University of the People

Department of Health Science

Author: MD Arman Hassan

This manuscript is an independent narrative review prepared by the author.

## Abstract

Modern social media platforms have evolved into algorithmically managed environments that measure social approval, reward self-presentation, and provide validation through variable reinforcement schedules. This critical literature review examines the ways in which these digital systems may favorably amplify subclinical narcissistic tendencies—defined here as non-pathological but identifiable patterns of vulnerable narcissism characterized by hypersensitivity to social evaluation, chronic reassurance-seeking, and unstable self-esteem—particularly among individuals with insecure attachment histories. Drawing on Kohut's self-psychology, Kernberg's object relations theory, Bowlby's attachment framework, and contemporary trait research on the grandiose-vulnerable distinction, the present review attempts to synthesize findings across three overlapping domains: vulnerable narcissism and social media behavior, mechanisms of algorithmic reinforcement, and attachment-related digital vulnerability. The reviewed literature suggests that vulnerable narcissists may engage with social media primarily for ego-protection rather than ego-enhancement, using controlled and selective self-presentation to regulate self-worth and avoid uncontrolled criticism. A conceptual tripartite model is proposed linking developmental attachment insecurity, individual-level narcissistic vulnerability, and contextual algorithmic reinforcement through a bidirectional feedback loop. Four testable propositions are derived to guide future longitudinal and cross-cultural research. Clinical and societal implications are discussed.

**Keywords:** *subclinical narcissism, vulnerable narcissism, algorithmic validation, ego-protection, social media reinforcement, attachment insecurity, digital mental health, selfobject mirroring*

## 1. Introduction

The question of whether social media makes people more narcissistic is widely discussed in both popular culture and academic research. The present review suggests that the question itself may be imprecise. A more useful question is whether algorithmically curated digital environments create conditions that selectively amplify pre-existing narcissistic vulnerabilities—and whether individuals with weaker early developmental histories may be most at risk. Understood this way, the relationship between narcissism and social media is not simply a story of technology corrupting character. It may be better understood as a story about the interaction between personality and environment, with significant implications for mental health.

Central to this analysis is the distinction between grandiose and vulnerable narcissism. Grandiose narcissism—characterized by entitlement, exhibitionism, and interpersonal dominance—has received disproportionate research attention, partly because the most widely used instrument, the Narcissistic Personality Inventory (NPI; Raskin & Terry, 1988), was designed to capture it. Vulnerable narcissism, by contrast, involves hypersensitivity to evaluation, shame-proneness, and a persistent oscillation between inflated and deflated self-states (Pincus & Lukowitsky, 2010). It is less visible, more clinically common, and—the present review proposes—may be more structurally suited to the psychological affordances of algorithm-driven platforms. Where grandiose narcissists appear to use social media to perform superiority and attract admiration, vulnerable narcissists may use it primarily for ego-protection: to manage fragile self-worth in a controllable environment and avoid unregulated criticism that face-to-face interaction cannot shield them from (Balcerowska et al., 2023). This distinction forms the basis of the analysis that follows.

Despite a growing body of research on narcissism and social media, theoretical integration remains limited. Most prior frameworks have not consistently distinguished between narcissistic subtypes, and the developmental roots of narcissistic vulnerability—specifically the role of insecure attachment—have rarely been systematically connected to digital reinforcement models (Casale et al., 2020). The present review attempts to address these gaps through three objectives: (1) to synthesize and critically evaluate evidence across three overlapping domains; (2) to identify methodological and theoretical limitations in current research; and (3) to propose a conceptual tripartite model suggesting how attachment insecurity, narcissistic vulnerability, and algorithmic reinforcement may interact in a self-perpetuating cycle.

## 2. Theoretical Foundations

### 2.1 Self-Psychology: The Mirroring Deficit

Heinz Kohut (1977) proposed that healthy psychological development depends substantially on the quality of early mirroring experiences—the degree to which caregivers accurately reflect and affirm the child's emerging sense of self. Where these experiences are chronically insufficient, the self may remain dependent on external affirmation as a primary regulatory mechanism. Within this framework, vulnerable narcissism is understood not as excess self-love but as

compensatory self-stabilization: the adult continuation of an unmet developmental need (Ronningstam, 2005).

A notable parallel may exist between Kohut's mirroring selfobject and algorithmic validation. Social media platforms provide quantified, visible, and continuous affirmation that resembles mirroring in form: they respond to self-presentation with approval metrics that individuals may rapidly incorporate into self-regulatory processes (Banai et al., 2005). An important limitation of this analogy should, however, be acknowledged. Kohut's mirroring selfobject was fundamentally a human relational phenomenon—the empathic attunement of a responsive other who perceives and reflects authentic affective states. Algorithmic feedback may be structurally similar in form but categorically different in nature: it responds not to the self's authentic states but to the self's performed presentations, through mechanisms indifferent to the individual's actual psychology. Whether this form of pseudo-mirroring produces the same regulatory effects as genuine selfobject experience—or whether it may generate a qualitatively different and potentially more destabilizing form of dependence—remains an empirical question the current literature has not yet adequately addressed.

### 2.2 Object Relations Theory: Splitting and Identity Instability

Kernberg's (1975) account of narcissism offers a complementary perspective that diverges from Kohut's in an important way. Where Kohut emphasized developmental deficit—the absence of adequate empathic provision—Kernberg emphasized structural conflict and the failure to integrate positive and negative self-representations into coherent wholes. This failure, which Kernberg termed splitting, results in the individual oscillating between idealized and devalued self-states depending on the nature of incoming social feedback.

Digital environments may sustain rather than resolve this splitting tendency. The binary logic of engagement metrics—the presence or absence of likes, the gain or loss of followers—provides continuous external inputs that may substitute for the internal integration Kernberg identifies as a key developmental achievement. A post that receives strong engagement may consolidate an idealized self-state; silence or criticism may activate a devalued one. Unlike a clinical therapeutic context, which aims to bring split representations into confrontation and promote integration, the social media environment appears to provide no such containing function. This is where Kernberg's framework may add something Kohut's does not: the suggestion that digital engagement is not merely an inadequate substitute for mirroring, but may actively reinforce the defensive structure that prevents genuine psychological integration.

### 2.3 Attachment Theory: Developmental Roots of Digital Vulnerability

Bowlby (1969) proposed that early caregiving relationships generate internal working models—cognitive-affective schemas that shape expectations about social availability and self-worth throughout the lifespan. Insecure attachment, particularly the anxious subtype, is characterized by hyperactivation of the attachment system: heightened vigilance to signs of rejection, compulsive reassurance-seeking, and emotional dysregulation in the face of perceived

unavailability (Mikulincer & Shaver, 2007). A substantial body of evidence suggests links between these patterns and vulnerable narcissism. Otway and Vignoles (2006) demonstrated quantitatively that attachment insecurity in childhood predicted narcissistic trait expression in adulthood, and Smolewska and Dion (2005) confirmed this association across multiple attachment dimensions.

Particularly relevant to the present review, a 2024 meta-analysis of 45 studies (N = 11,746) found that both anxious and avoidant attachment significantly predicted problematic social media use, with anxious attachment showing the strongest and most consistent effect across Western and Eastern samples (r = .31, 95% CI [.26, .36]; Zhang et al., 2024a). This finding appears to provide meaningful empirical grounding for the developmental level of the proposed conceptual model and suggests that digital environments may represent a contemporary context in which attachment-derived reassurance-seeking is expressed and potentially reinforced algorithmically.

### 2.4 Contemporary Trait Research: Why the Grandiose-Vulnerable Distinction Matters

Contemporary personality research has established that grandiose and vulnerable narcissism are empirically distinct constructs with different motivational structures and behavioral outcomes (Wink, 1991; Pincus & Lukowitsky, 2010). This distinction appears to have direct implications for the social media literature. Grandiose narcissists seem to approach social media with an ego-enhancement motive—seeking admiration and broadcasting status. Vulnerable narcissists, by contrast, may approach it with an ego-protection motive —seeking to stabilize fragile self-worth through controlled self-presentation and the avoidance of unregulated criticism.

Balcerowska et al. (2023) empirically explored this distinction across two culturally diverse samples (N = 1,946 Polish and German participants), finding that negative reinforcement expectancies—the anticipated relief from negative affect—were the stronger predictor of social media addiction for vulnerable narcissism ($\beta = .41$, $p < .001$), while positive reinforcement expectancies were the stronger predictor for grandiose narcissism ($\beta = .38$, $p < .001$). Many prior integrative frameworks did not fully anticipate this motivational difference, and it appears to have meaningful implications for both theoretical modeling and intervention design.

## 3. Methodology

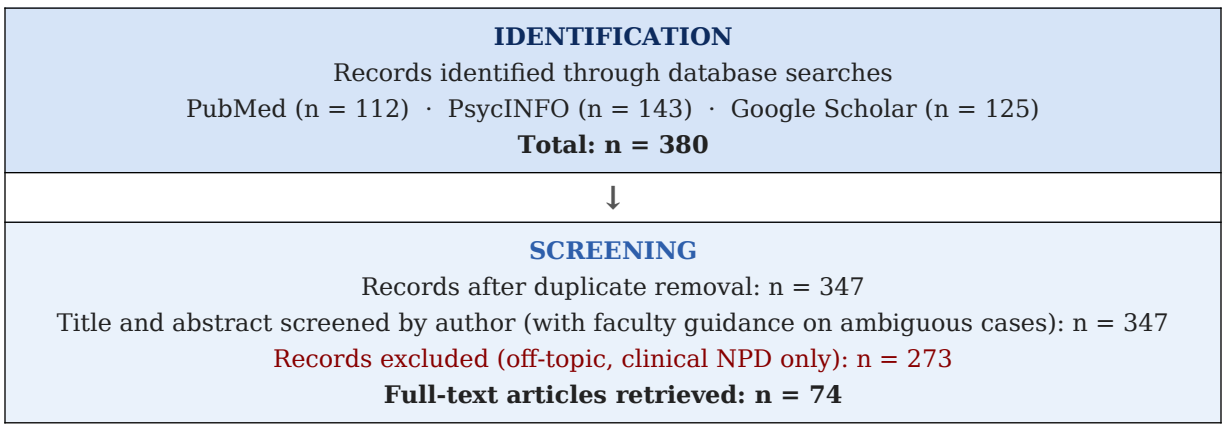
This review employs a narrative critical review methodology (Grant & Booth, 2009), which synthesizes theoretical and empirical literature across interdisciplinary domains while evaluating the strengths, weaknesses, and conceptual coherence of included studies. As a narrative rather than systematic review, the synthesis was guided by theoretical relevance and conceptual fit rather than exhaustive coverage. The critical component involved moving beyond descriptive summary to assess where evidence appears strong, where it is constrained by methodological design, and where theoretical claims may exceed available empirical support.

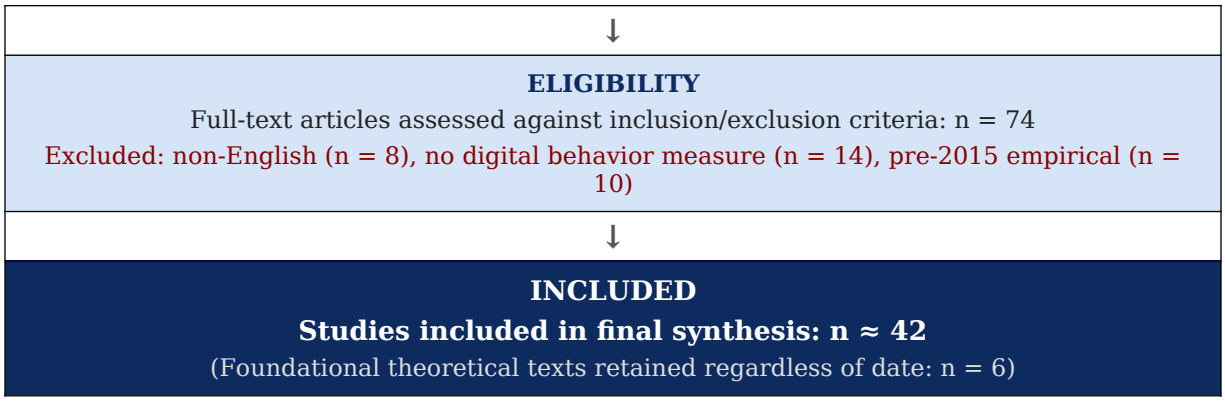
Literature was searched by the author across PubMed, PsycINFO, and Google Scholar, with faculty guidance during the selection process. The following Boolean search strings were used: for PubMed and PsycINFO: ("subclinical narcissism" OR "vulnerable narcissism" OR "narcissistic traits") AND ("social media" OR "digital platforms" OR "algorithmic feedback") AND ("mental health" OR "self-esteem" OR "attachment"); and for Google Scholar: "vulnerable narcissism AND social media reinforcement," "narcissistic traits AND attachment insecurity AND digital," and "algorithmic validation AND self-esteem." Searches were restricted to peer-reviewed articles in English published between January 2015 and June 2025. Foundational theoretical texts (Bowlby, 1969; Kohut, 1977; Kernberg, 1975) were retained regardless of publication date given their theoretical centrality. Studies were included if they addressed narcissistic traits, digital platform behavior, or attachment processes in adolescent or adult populations. Clinical-only NPD studies without reference to subclinical dimensions were excluded.

An initial search yielded approximately 380 records. Following title and abstract screening by the author, with faculty guidance on ambiguous cases (initial agreement rate approximately 89%; discrepancies resolved by discussion), 74 full-text articles were retrieved for closer review. These 74 articles were assessed against predetermined inclusion and exclusion criteria, resulting in approximately 42 articles considered most relevant for synthesis. A PRISMA-inspired flow diagram illustrating the search and selection process is presented in Figure 2. Articles were organized using a thematic approach informed by Thomas and Harden (2008): each article was read in full, then coded for theoretical framework, methodology, population, and key findings. Codes were progressively grouped into four thematic domains—narcissistic subtypes and platform behavior, theoretical frameworks and digital applicability, algorithmic reinforcement mechanisms, and attachment pathways—which informed both the review's structure and the proposed conceptual model. As a further quality check, studies were informally appraised on three criteria: methodological transparency, sample diversity, and theoretical contribution. Studies that appeared weak on multiple criteria were treated as supplementary rather than primary evidence throughout the discussion.

***Figure 1***

*PRISMA-Inspired Flow Diagram of Literature Search and Selection Process*


IDENTIFICATION
Records identified through database searches
PubMed (n = 112) · PsycINFO (n = 143) · Google Scholar (n = 125)
Total: n = 380
↓
SCREENING
Records after duplicate removal: n = 347
Title and abstract screened by author (with faculty guidance on ambiguous cases): n = 347
Records excluded (off-topic, clinical NPD only): n = 273
Full-text articles retrieved: n = 74


↓
ELIGIBILITY
Full-text articles assessed against inclusion/exclusion criteria: n = 74
Excluded: non-English (n = 8), no digital behavior measure (n = 14), pre-2015 empirical (n = 10)
↓
INCLUDED
Studies included in final synthesis: n ≈ 42
(Foundational theoretical texts retained regardless of date: n = 6)

*Note. Literature was searched by the author with faculty guidance during the selection process. Foundational texts (Bowlby, 1969; Kohut, 1977; Kernberg, 1975; Ferster & Skinner, 1957; Fogg, 2003; Mikulincer & Shaver, 2007) were retained regardless of publication date given their theoretical centrality.*

## 4. Persistent Limitations in Current Research

Three interconnected limitations recur across the existing literature. Together, they may help explain why no current framework appears to have produced a fully adequate account of the narcissism-algorithm-attachment relationship, and they provide part of the rationale for the conceptual model proposed in Section 8.

### 4.1 Measurement Inconsistency and Subtype Conflation

The NPI (Raskin & Terry, 1988) remains dominant in the social media literature despite its well-documented grandiosity bias. Studies relying exclusively on the NPI, including several high-profile meta-analyses (e.g., McCain & Campbell, 2018), may systematically underrepresent vulnerable narcissism. The Pathological Narcissism Inventory (PNI; Pincus et al., 2009) and the Hypersensitive Narcissism Scale (Hendin & Cheek, 1997) address this gap but are used inconsistently, limiting cross-study comparability. Critically, most social media studies have not distinguished between subtypes in their designs, even though grandiose and vulnerable narcissism appear to predict digital behavior through meaningfully different mechanisms—ego-boosting versus ego-protection respectively (Balcerowska et al., 2023).

### 4.2 The Directionality Problem

The dominant research design is cross-sectional and self-report, which cannot resolve whether vulnerable narcissistic traits may predispose individuals toward problematic digital engagement, or whether intensive platform use might amplify narcissistic traits over time, or whether both processes operate simultaneously. The reinforcing spirals model (Slater, 2007) suggests that both directions of effect may be active, and some longitudinal evidence is consistent with this possibility. Halpern et al. (2016), using a cross-lagged panel design, found mutual amplification between narcissism and selfie production over one year —though this study focused on grandiose narcissism only. The absence of

longitudinal evidence specifically addressing vulnerable narcissism remains a notable empirical gap in the field.

### 4.3 Cultural Underrepresentation

The overwhelming majority of included studies were conducted with Western, educated, industrialized, rich, and democratic (WEIRD) samples. Constructs central to this review—entitlement, public approval-seeking, impression management—likely carry different normative meanings across collectivistic and individualistic cultural contexts (Twenge & Campbell, 2019). A notable exception is Fioravanti et al. (2025), whose cross-cultural network analysis of TikTok users across China, Italy, and Spain (N = 3,313) found both shared mechanisms and culturally specific expressions of digital vulnerability. The 2024 meta-analysis also found that Eastern samples showed somewhat stronger attachment-digital vulnerability associations than Western samples (Zhang et al., 2024a), suggesting that cultural context may meaningfully moderate the mechanisms discussed in this review.

## 5. Vulnerable Narcissism and Social Media: The Ego-Protection Dynamic

A central argument of the present review—one that attempts to distinguish it from earlier frameworks—is that vulnerable narcissists may engage with social media primarily through an ego-protection rather than ego-enhancement motive. This distinction appears to have meaningful theoretical and practical implications.

Grandiose narcissists appear drawn to social media as a performance stage: they seek admiration, broadcast superiority, and extract positive feedback that may confirm their inflated self-concept (McCain & Campbell, 2018). Vulnerable narcissists, by contrast, may enter digital environments carrying a deficit—fragile self-worth, hypersensitivity to criticism, and a motivation to avoid the ego-threatening encounters that can characterize offline social life. Social media may offer them what offline environments often cannot: controlled self-presentation, selective disclosure, the ability to avoid or delete negative feedback, and curated audiences (Brailovskaia et al., 2023). This protection motive was examined by Balcerowska et al. (2023), who found it to be the dominant predictor of social media addiction for vulnerable narcissism across two national samples.

Within this protective dynamic, quantifiable feedback—likes, comments, and follower counts—may take on amplified psychological significance. For vulnerable narcissists, these metrics may function as temporary self-worth stabilizers. High engagement appears to provide short-lived relief from self-doubt; low engagement or negative comments may trigger acute self-esteem crashes (Casale et al., 2021; Tully, 2024). Social comparison processes may further intensify this dynamic: vulnerable narcissism has been associated with both upward and downward social comparisons on social networking sites, with upward comparisons predicting greater self-esteem instability and downward comparisons providing only temporary relief (Kong et al., 2021). In a study of N = 847 participants, the oversensitivity dimension of vulnerable narcissism emerged as the strongest predictor of all six measured aspects of platform use—including time spent,

posting frequency, and preoccupation with likes ($r = .38–.51$, all $p < .001$)—while the egocentricity dimension showed a more complex protective pattern (Casale et al., 2021). These findings appear consistent with the model's individual level: oversensitivity may be a particularly important driver of compulsive platform engagement among this population.

There may also be a paradox at the center of this dynamic. The more successfully a vulnerable narcissist manages their digital identity to attract approval, the more that approval may be experienced as conditional on the performance rather than confirming of the authentic self—and therefore the less effective it becomes as a genuine self-regulatory resource. Drawing on Kohut's framework, transmuting internalization may not occur when the selfobject function is provided by a system that responds to performance rather than authentic self-expression. This could be one mechanism through which ego-protection becomes self-perpetuating over time.

## 6. Algorithmic Architecture as Psychological Infrastructure

Social media platforms are not psychologically neutral environments. Research in persuasive technology and behavioral design suggests that they are structured to maximize engagement through variable reward schedules—mechanisms known since Ferster and Skinner (1957) to produce some of the highest rates of behavioral persistence of any reinforcement pattern. Unlike fixed schedules that allow individuals to predict when rewards will arrive, variable schedules maintain persistent engagement through unpredictability. For individuals with vulnerable narcissism whose self-esteem may fluctuate in response to social feedback, the unpredictability of digital validation could be particularly reinforcing: it may mirror their internal regulatory experience, where feedback sometimes stabilizes self-worth and sometimes destabilizes it, driving compulsive monitoring (Fogg, 2003; Alter, 2017).

A 2025 systematic review analyzing a decade of research across 30 studies found that algorithmic systems appear to structurally amplify selective social exposure and feedback patterns, with effects most pronounced among youth populations (Hamari et al., 2025). Importantly, these effects may not be uniform across users. Research suggests that higher levels of psychological vulnerability—including low self-esteem and attachment anxiety—predict more problematic engagement with these reinforcement structures (Zhang et al., 2024a; Fioravanti et al., 2025). This conditional susceptibility is central to the proposed conceptual framework: algorithms may create a reinforcement infrastructure, but it appears to be the individual's developmental and personality profile that determines how deeply that infrastructure is internalized.

The quantification of social approval adds a further dimension with limited precedent in classical psychological frameworks. Likes, follower counts, and engagement rates transform what was previously an ambiguous and largely unverifiable experience of social standing into a precise, public, and continuously updated numerical index. For individuals with vulnerable narcissism, whose self-evaluation already appears hyperresponsive to social cues, this quantification may represent a qualitative transformation of the social environment. Ambiguity—

which is necessary for psychological integration—is reduced. The self-state activated by a given interaction may become less a matter of interpretation and more a matter of arithmetic.

## 7. Attachment Pathways in Digital Contexts

The developmental mechanism proposed in this review is cyclical in nature. Individuals with anxious attachment may enter digital environments primed for reassurance-seeking (Mikulincer & Shaver, 2007). The algorithmic architecture of platforms like TikTok and Instagram rewards consistent posting, emotional content, and social engagement—potentially incentivizing the validation-seeking behavior that anxiously attached individuals may be motivationally predisposed toward. Initial positive engagement may reduce anxiety temporarily; inconsistent or negative responses could amplify it, driving increased engagement to restore a temporary equilibrium. A path analysis model of 529 college students found that attachment insecurity predicted fear of missing out (FoMO) and reduced self-esteem, which in turn predicted problematic social media use—a finding consistent with the sequential mechanism proposed here (Monacis et al., 2023). Additionally, a three-level meta-analysis found that attachment insecurity—both anxious and avoidant—was a consistent predictor of pathological narcissism across developmental stages (Zhang et al., 2024b), which appears to support the developmental level of the proposed model. Over time, this cycle may consolidate digital platforms as primary sources of soothing that are simultaneously unreliable and compulsively sought.

A 2025 cross-cultural network analysis found that attachment anxiety was a central node in the psychological vulnerability network predicting problematic TikTok use across Chinese, Italian, and Spanish adolescents (Fioravanti et al., 2025)—a finding that may offer preliminary multi-national support for this pathway. The 2024 meta-analysis also found that the attachment-digital vulnerability association appeared somewhat stronger in Eastern samples than in Western samples (Zhang et al., 2024a), suggesting that cultural context may moderate the relationship in ways the present model does not fully capture.

It is important to be clear about what the current evidence does not yet establish. The meta-analytic findings are correlational, and the longitudinal evidence needed to confirm that attachment insecurity causally predicts digital reinforcement cycles remains very limited. This is among the most important empirical gaps that the four propositions of this review are designed to highlight.

## 8. A Proposed Integrative Tripartite Model

The present review proposes a conceptual framework that attempts to connect—rather than definitively resolve—the relationship among attachment insecurity, narcissistic vulnerability, and algorithmic reinforcement. The goal is not to present a finished theory, but to offer a preliminary model that integrates prior frameworks and specifies directional pathways worth testing empirically. Few existing single-domain theories appear to address the direction, sequence, and

reciprocal nature of these interactions simultaneously. That gap is the motivation for the model presented here (see Figure 1).

***Figure 2***

*The Integrative Tripartite Model of Algorithmic Narcissistic Reinforcement*

LEVEL 1 — Developmental
Attachment Insecurity
Early caregiving deficits → internal working models → chronic external reassurance-seeking
LEVEL 2 — Individual
Narcissistic Vulnerability
Hypersensitivity to evaluation · ego-protection motive · self-esteem instability
LEVEL 3 — Contextual
Algorithmic Reinforcement
Variable reward schedules · quantified approval metrics · ego-protective avoidance loop
Bidirectional feedback loop

*Note. Arrows indicate proposed direction of effect. The feedback loop represents the model's central contribution: contextual reinforcement cycles may progressively deepen the developmental and individual-level vulnerabilities from which they emerge.*

At the developmental level, early attachment insecurity may generate internal working models characterized by chronic reassurance-seeking and heightened sensitivity to social rejection (Bowlby, 1969; Mikulincer & Shaver, 2007). This level is proposed to establish a predispositional substrate—a vulnerability that may make the individual differentially responsive to what the contextual level provides. It may also help explain why individuals with more secure attachment histories appear to engage with the same digital environments from a position of relative regulatory stability.

At the individual level, vulnerable narcissistic traits—particularly the oversensitivity dimension, which Casale et al. (2021) found to be an important driver of compulsive platform engagement—may constitute the personality-level expression of unresolved attachment insecurity (Otway & Vignoles, 2006). This level is proposed to bridge developmental history and contextual exposure by suggesting the motivational structure through which digital environments are engaged: primarily avoidance motivation directed toward managing self-worth and minimizing ego threat, rather than approach motivation directed toward admiration (Balcerowska et al., 2023).

At the contextual level, algorithmically governed platforms may provide a reinforcement architecture that appears structurally matched to the motivational profile of vulnerable narcissism. Variable reward schedules may sustain compulsive engagement; quantified approval metrics may reduce the ambiguity of social feedback; curated self-presentation may afford the ego-protective identity management that vulnerable narcissism motivates. Importantly, this contextual impact is not proposed to be uniform—it may be experienced most strongly by those with the vulnerabilities described at the developmental and individual levels.

The feedback loop—the model's proposed central contribution—is suggested to operate as follows: sustained engagement with algorithmically variable social feedback may not merely maintain existing narcissistic traits, but could progressively deepen the attachment insecurity from which those traits appear to derive. Each cycle of validation-seeking and pseudo-mirroring may reinforce the behavioral pattern without resolving the underlying regulatory deficit. This generates a preliminary, theoretically derived prediction: the effects of digital engagement on narcissistic vulnerability may increase over time rather than stabilize—a claim no existing cross-sectional study is positioned to evaluate.

Four preliminary testable propositions follow from this framework:

**Proposition 1**: Anxious attachment may moderate the relationship between social media feedback variability and self-esteem instability, such that anxiously attached individuals could show disproportionate self-worth fluctuation in response to variable engagement metrics, above and beyond the independent contribution of vulnerable narcissistic traits.

**Proposition 2**: Vulnerable narcissism, compared to grandiose narcissism, may show a stronger association with ego-protective digital behaviors—including avoidance of negative feedback, deletion of low-engagement content, and preferential engagement with affirming audiences—potentially mediated by negative reinforcement expectancies.

**Proposition 3**: Cultural collectivism may moderate the attachment insecurity-digital validation cycle, such that the association between anxious attachment and problematic social media use could appear stronger in individualistic cultural contexts, where the personal salience of public approval metrics may be amplified by individual achievement norms.

**Proposition 4**: Prolonged exposure to algorithmically variable social feedback may independently predict longitudinal increases in both vulnerable narcissistic trait expression and anxious attachment orientation, above and beyond baseline measures—a finding that would be consistent with the model's suggestion that digital reinforcement cycles deepen rather than merely reflect pre-existing vulnerabilities. These propositions are offered as preliminary hypotheses to guide future empirical work, not as established findings.

## 9. Clinical and Societal Implications

### 9.1 Emotional Regulation and Identity

The proposed model may carry some practical implications for clinical contexts. It suggests that addressing social media behavior in isolation—through usage monitoring or platform restriction alone—could produce limited and temporary benefits if the underlying attachment and personality-level vulnerabilities remain unaddressed. Approaches that more directly target internal working models and self-regulatory deficits, such as mentalization-based therapy (Bateman & Fonagy, 2016) or schema-focused interventions, may be better positioned to address the cycle at a more foundational level. These suggestions are

speculative and would require empirical validation before informing clinical practice.

The model also suggests that self-regulatory deterioration may occur gradually and without subjective awareness, through repeated cycles of pseudo-mirroring that appear to provide relief while potentially eroding internal regulation over time. If this pattern holds, it may have some practical relevance for how clinicians approach social media use: not merely as a behavioral pattern to quantify, but as a possible indicator of the patient's self-regulatory functioning and the degree to which external digital systems may be substituting for autonomous affect regulation.

### 9.2 Interpersonal Functioning

The ego-protection dynamic may also carry interpersonal consequences that extend beyond the individual. Vulnerable narcissists who regulate self-worth primarily through curated digital presentation could, over time, develop reduced tolerance for the ambiguous and less controllable evaluative encounters that characterize many authentic interpersonal relationships. If digital environments provide more manageable—though potentially less nourishing—social feedback than face-to-face relationships, heavy sustained engagement might affect the motivational and regulatory capacities involved in genuine intimacy. This is not a technological-deterministic claim; it is a psychological hypothesis about what highly controlled environments may structurally fail to provide.

### 9.3 Preventive Implications

At the population level, the proposed model suggests that universal digital literacy interventions—while potentially helpful for many users—may not be sufficient for individuals whose digital risk is rooted in structural motivational vulnerability rather than inadequate knowledge. Prevention efforts directed at higher-risk individuals might benefit from also addressing attachment security and self-regulatory capacity. The developmental level of the model implies that the most effective preventive window may precede heavy social media use: in the quality of early caregiving, in educational environments that can support the development of internal self-evaluation, and in relational contexts that foster secure attachment. These implications remain theoretical and would need to be tested empirically before informing policy or practice.

## 10. Future Research Directions

The four propositions derived from the proposed model suggest a research agenda that prioritizes designs capable of addressing directional and mechanistic questions that cross-sectional research cannot resolve.

Longitudinal designs appear to be the most pressing need. Multi-wave panel studies assessing attachment orientation, narcissistic subtype, digital engagement patterns, and self-esteem trajectories across six months to two years would be well positioned to test Proposition 4. Ecological momentary assessment (EMA) approaches, which capture self-esteem fluctuations in real time in response to

discrete social media interactions, could allow the mechanisms in Propositions 1 and 2 to be examined with temporal precision that retrospective self-report cannot achieve.

Experimental manipulation of feedback variability represents a second area of priority. Designs that randomly assign participants to consistent versus variable social media feedback conditions—while measuring self-esteem trajectory, attachment activation, and behavioral persistence—could provide causal evidence that observational research is unable to generate. This approach appears technically feasible but would raise ethical considerations requiring careful institutional review.

Cross-cultural and platform-comparative studies represent a third important direction. Testing Proposition 3 would require cross-national samples with comparable measurement of both attachment orientation and narcissistic subtype. Studies comparing platforms with different algorithmic architectures—Instagram's image-focused feedback versus TikTok's viewing-duration algorithm versus Twitter's controversy-rewarding engagement—could allow more precise identification of which algorithmic features may be most relevant to vulnerable narcissistic engagement patterns.

## 11. Conclusion

The present review has attempted to argue that subclinical vulnerable narcissism, attachment insecurity, and algorithmic reinforcement may constitute three interacting levels of a self-sustaining vulnerability cycle—one that existing single-domain frameworks have not fully addressed. The tripartite framework proposed here attempts to integrate three theoretical connections: that ego-protection rather than ego-enhancement may be the primary motivational mechanism linking vulnerable narcissism to digital engagement; that a bidirectional feedback loop could cause contextual reinforcement to deepen rather than merely reflect developmental and personality-level vulnerabilities; and that both mechanisms may be grounded in a developmental theory of attachment that helps explain differential susceptibility. These remain theoretical proposals, and their empirical status depends on future longitudinal and experimental research.

The proposed model also attempts to clarify something the social media and narcissism debate has sometimes obscured: algorithms may not generate narcissistic vulnerability from nothing. They may amplify vulnerabilities that already exist, in individuals whose developmental histories have made them structurally susceptible to the particular form of pseudo-mirroring that variable reinforcement schedules appear to provide. If this interpretation is supported by future research, it would suggest that addressing these phenomena meaningfully requires developmental and relational responses, not merely digital ones.

## References

Alter, A. (2017). *Irresistible: The rise of addictive technology and the business of keeping us hooked.* Penguin.

Balcerowska, J. M., Bereznowski, P., Biernatowska, A., Atroszko, P. A., Sawicki, P., & Bereznowski, P. (2023). Different aspects of narcissism and social networking sites addiction in Poland and Germany: The mediating role of positive and negative reinforcement expectancies. *Personality and Individual Differences, 205,* 111719. https://doi.org/10.1016/j.paid.2023.111719

Banai, E., Mikulincer, M., & Shaver, P. R. (2005). "Selfobject" needs in Kohut's self psychology: Links with attachment, self-cohesion, affect regulation, and adjustment. *Psychoanalytic Psychology, 22*(2), 224–260. https://doi.org/10.1037/0736-9735.22.2.224

Bateman, A., & Fonagy, P. (2016). *Mentalization-based treatment for personality disorders: A practical guide.* Oxford University Press.

Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment.* Basic Books.

Brailovskaia, J., Teismann, T., Margraf, J., & Cosci, F. (2023). Vulnerable narcissism, fear of missing out (FoMO), and addictive social media use: A gender comparison from Germany. *Computers in Human Behavior, 141,* 107620. https://doi.org/10.1016/j.chb.2023.107620

Casale, S., Fioravanti, G., Rugai, L., & Flett, G. L. (2020). Narcissism and problematic social media use: A systematic literature review. *Computers in Human Behavior Reports, 2,* 100032. https://doi.org/10.1016/j.chbr.2020.100032

Casale, S., Rugai, L., & Fioravanti, G. (2021). Social media use and vulnerable narcissism: The differential roles of oversensitivity and egocentricity. *International Journal of Environmental Research and Public Health, 18*(17), 9125. https://doi.org/10.3390/ijerph18179125

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement.* Appleton-Century-Crofts.

Fioravanti, G., Cerolini, S., Svicher, A., & Casale, S. (2025). Examining the psychological vulnerabilities and motives behind problematic TikTok use among youths: A cross-cultural study using a network approach. *International Journal of Human-Computer Interaction, 42*(1), 509–524. https://doi.org/10.1080/10447318.2025.2508312

Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do.* Morgan Kaufmann.

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Halpern, D., Valenzuela, S., & Katz, J. E. (2016). "Selfie-ists" or "narci-selfiers"?: A cross-lagged panel analysis of selfie taking and narcissism. *Personality and Individual Differences, 97,* 98–101. https://doi.org/10.1016/j.paid.2016.03.019

Hamari, J., Bujić, M., Macey, A. L., & Kerous, B. (2025). Trap of social media algorithms: A systematic review of research on filter bubbles, echo

chambers, and their impact on youth. *Social Sciences, 15*(11), 301. https://doi.org/10.3390/socsci15110301

Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: A reexamination of Murray's Narcism Scale. *Journal of Research in Personality, 31*(4), 588–599. https://doi.org/10.1006/jrpe.1997.2204

Kernberg, O. F. (1975). *Borderline conditions and pathological narcissism.* Jason Aronson.

Kohut, H. (1977). *The restoration of the self.* International Universities Press.

Kong, F., Zhong, M., Ding, K., & Chen, H. (2021). Vulnerable narcissism in social networking sites: The role of upward and downward social comparisons. *Frontiers in Psychology, 12,* 711909. https://doi.org/10.3389/fpsyg.2021.711909

McCain, J. L., & Campbell, W. K. (2018). Narcissism and social media use: A meta-analytic review. *Psychology of Popular Media Culture, 7*(3), 308–327. https://doi.org/10.1037/ppm0000137

Mikulincer, M., & Shaver, P. R. (2007). *Attachment in adulthood: Structure, dynamics, and change.* Guilford Press.

Monacis, L., de Palo, V., Griffiths, M. D., & Sinatra, M. (2023). The associations between attachment, self-esteem, fear of missing out, daily time expenditure, and problematic social media use: A path analysis model. *Computers in Human Behavior, 141,* 107629. https://doi.org/10.1016/j.chb.2023.107629

Otway, L. J., & Vignoles, V. L. (2006). Narcissism and childhood recollections: A quantitative test of psychoanalytic predictions. *Personality and Social Psychology Bulletin, 32*(1), 104–116. https://doi.org/10.1177/0146167205279907

Pincus, A. L., & Lukowitsky, M. R. (2010). Pathological narcissism and narcissistic personality disorder. *Annual Review of Clinical Psychology, 6,* 421–446. https://doi.org/10.1146/annurev.clinpsy.121208.131215

Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G. C., & Levy, K. N. (2009). Initial construction and validation of the Pathological Narcissism Inventory. *Psychological Assessment, 21*(3), 365–379. https://doi.org/10.1037/a0016530

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54*(5), 890–902. https://doi.org/10.1037/0022-3514.54.5.890

Ronningstam, E. (2005). *Identifying and understanding the narcissistic personality.* Oxford University Press.

Slater, M. D. (2007). Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social

identity. *Communication Theory, 17*(3), 281–303. https://doi.org/10.1111/j.1468-2885.2007.00296.x

Smolewska, K., & Dion, K. (2005). Narcissism and adult attachment: A multivariate approach. *Self and Identity, 4*(1), 59–68. https://doi.org/10.1080/13576500444000218

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology, 8*(1), 45. https://doi.org/10.1186/1471-2288-8-45

Tully, K. (2024). *The role of narcissistic personality traits in the affective experience of using Instagram* (Doctoral dissertation). Long Island University. https://digitalcommons.liu.edu/brooklyn_fulltext_dis/43

Twenge, J. M., & Campbell, W. K. (2019). Media use is linked to lower psychological well-being: Evidence from three datasets. *Psychiatric Quarterly, 90*(2), 311–331. https://doi.org/10.1007/s11126-019-09630-7

Wink, P. (1991). Two faces of narcissism. *Journal of Personality and Social Psychology, 61*(4), 590–597. https://doi.org/10.1037/0022-3514.61.4.590

Zhang, R., Li, Y., Liu, B., & Chen, X. (2024b). The relationship between attachment insecurity and pathological narcissism: A three-level meta-analysis. *Journal of Family Theory & Review.* https://doi.org/10.1111/jftr.12593

Zhang, Y., Chen, M., & Wang, L. (2024a). Adult attachment, social anxiety, and problematic social media use: A meta-analysis and meta-analytic structural equation model. *Addictive Behaviors, 156,* 107998. https://doi.org/10.1016/j.addbeh.2024.107998

## Author Note

The author is an university undergraduate student of the Department of Health Science in University of the People. The present manuscript is created in the form of a separate academic review. Grammar correction was done by using language editing tools alone. There were no external funds obtained. The author does not declare any conflict of interest.